УДК 519.24

СТАТИСТИЧЕСКАЯ ОЦЕНКА ВЕСОВОГО КОЭФФИЦИЕНТА В КОМБИНИРОВАННОЙ МОДЕЛИ РЕГРЕССИИ

С.В. Скрипин

Томский государственный университет
Томский научный центр СО РАН
E-mail: skripin@ef.tsu.ru

Представлены варианты статистической оценки весового коэффициента в комбинированной модели регрессии, использующей оценки построенных моделей – параметрической и непараметрической (в задачах регрессии или дискриминации). Вариант модифицированной статистической оценки весового коэффициента позволяет получать оценки регрессии в заданной точке, близкие к оптимальным по критерию минимума среднеквадратического отклонения. Результаты статистического моделирования на выборках конечного объема показывают, что комбинированная оценка регрессии со статистической оценкой весового коэффициента предпочтительнее, чем каждая из оценок построенных моделей.

Ключевые слова:
Статистическая оценка, комбинированная оценка, непараметрическая регрессия, статистическое моделирование.

Key words:
Statistical estimate, combined estimate, non-parametric regression, statistical simulation.

Введение

Задача оценки регрессии в условиях конечных выборок малого объема может быть решена с использованием различных моделей, как из класса параметрических, так и непараметрических. Для повышения качества оценок в указанных условиях выгоднее использовать комбинированные модели регрессии.

Одним из перспективных подходов для оценки регрессии является получение оценок комбинированных моделей, использующих оценки построенных моделей регрессии и априорную информацию о некоторых функционалах плотности [1]. В условиях отсутствия априорной информации о выборке конечного объема N качество оценок комбинированных моделей (для задач регрессии или дискриминации) может быть лучше, чем у моделей из класса параметрических или непараметрических [2, 3]. Результаты работы [1] позволяют оперировать множеством построенных моделей из разных классов и рассматривать в виде результатов классы оценок. В [1] класс оценок регрессии в точке $\mathbf{x}\in R^m$ представлен следующим выражением

$$\mathbf{J}_N(\mathbf{x};\boldsymbol{\lambda}) = \mathbf{J}_N(\mathbf{x}) - \boldsymbol{\lambda}^T\widehat{\Delta}(\mathbf{x}), \qquad (1)$$

где $\mathbf{J}_N(\mathbf{x})$ – непараметрические оценки неизвестной функции регрессии; $\boldsymbol{\lambda}$ – вектор коэффициентов, выбираемый согласно требованиям к качеству оценки; $\widehat{\Delta}(\mathbf{x})$ – матрица оценок, содержащая некоторую дополнительную информацию.

Практическое применение оценки (1) требует знания вектора $\boldsymbol{\lambda}$. Можно предложить различные методы его оценивания. В каждом методе можно использовать разные способы и критерии оценивания вектора. В результате получим разные по структуре выражения для $\boldsymbol{\lambda}$ и разные его свойства по способам оценивания, а следовательно, и свойства оценки (1).

Цель работы

Целью работы является предложить новый подход к оцениванию $\boldsymbol{\lambda}$ и сравнить при конечном N качество предложенных оценок с полученными ранее. С этой целью проведем имитационный эксперимент на упрощенной модели регрессии, построенной по оценкам двух моделей, например, параметрической и непараметрической [2, 4, 5]

$$J_N(\mathbf{x};\lambda) = \lambda J(\mathbf{x};\hat{\boldsymbol{\theta}}) + (1-\lambda)J_N(\mathbf{x}), \qquad (2)$$

где λ – скалярный весовой коэффициент; $J(\mathbf{x};\hat{\boldsymbol{\theta}})$ – параметрическая оценка регрессии; $\hat{\boldsymbol{\theta}}$ – оценка вектора $\boldsymbol{\theta}=(\theta_0,\ldots,\theta_m)$ по методу наименьших квадратов; $J_N(\mathbf{x})$ – непараметрическая оценка регрессии.

В [2] показано, что комбинированная оценка (2) по критерию минимума среднеквадратической ошибки (СКО) предпочтительнее оценок построенных моделей (параметрической или непараметрической) как в случае конечного объема выборок N, так и в асимптотическом случае. Наибольший практический интерес представляет случай конечного объема выборок. При подстановке различных оценок λ в (2) получим множество оценок $J_N(\mathbf{x};\lambda)$. Возникает проблема выбора оценок $J_N(\mathbf{x};\lambda)$ при конечном N, а следовательно – оценок λ. Основная проблема заключается в получении при конечном N наилучших коэффициентов λ по критерию минимума СКО оценок $J_N(\mathbf{x};\lambda)$. Поскольку истинная функция регрессии $J(\mathbf{x})$ неизвестна, то неизвестна и СКО.

В [2] предложено несколько способов оценки скалярного весового коэффициента λ для частного случая выражения (2). Для сравнения представим комбинированную модель с оценкой коэффициента λ, не зависящей от выбора точки $\mathbf{x}$ [2–4], показавшую в имитационном эксперименте наилучшие результаты.

Обозначим как $(\mathbf{X}_1, Y_1),...,(\mathbf{X}_N, Y_N)$ независимые наблюдения случайного вектора $(\mathbf{X}, Y)$ с неизвестной плотностью вероятности $f(\mathbf{x},y)$ (относительно меры Лебега); $\mathbf{x}\in R^m$, $y\in R^1$. Для построения комбинированных оценок выберем две модели регрессии из разных классов: параметрическую $J(\mathbf{x};\boldsymbol{\theta})$ и непараметрическую $J(\mathbf{x}) = \int_{-\infty}^{\infty} yf(y\,|\,\mathbf{x})dy$, где $f(y|\mathbf{x})$ – неизвестная условная плотность вероятности случайной величины Y при фиксированном значении $\mathbf{X}=\mathbf{x}$.

В качестве критерия оптимизации указанной оценки коэффициента λ выбрано выражение суммы квадратов регрессионных остатков (выборочный критерий).

$$Q_1(\lambda) = \sum_{i=1}^{N}[Y_i - J_N(\mathbf{X}_i;\lambda)]^2 \to \min_{\lambda},$$
$$\mathbf{X}_i = \{X_i^{(1)},...,X_i^{(m)}\}. \tag{3}$$

Из (3) с учетом (2) получена выборочная оценка оптимального коэффициента $\lambda=\lambda_o$:

$$\hat{\lambda}_o = \frac{\sum_{i=1}^{N}(J_N(\mathbf{X}_i) - Y_i)(J_N(\mathbf{X}_i) - J(\mathbf{X}_i;\hat{\boldsymbol{\theta}}))}{\sum_{i=1}^{N}(J_N(\mathbf{X}_i) - J(\mathbf{X}_i;\hat{\boldsymbol{\theta}}))^2}. \tag{4}$$

Оценка (4) коэффициента λ_o не зависит от выбора точки $\mathbf{x}\in R^m$, в которой оценивается регрессия, и требует малых затрат вычислительных ресурсов, что является преимуществом оценки λ_o. При соблюдении условий различий в оценках построенных моделей (параметрической и непараметрической) оценка (4) применима при конечном N. Но следует заметить, что коэффициент λ_o не гарантирует наилучших оценок регрессии в *заданной* точке $\mathbf{x}$, отличной от выборочных $\mathbf{X}_i$.

Постановка задачи

Для большинства задач прогнозирования, использующих модели регрессии, наибольший интерес представляет оценка регрессии в заданных точках $\mathbf{x}$, которые могут не совпадать с выборочными $\mathbf{X}_i$. Следовательно, необходима реализация комбинированной оценки регрессии с наилучшей оценкой коэффициента λ в заданной точке $\mathbf{x}$ – $\lambda(\mathbf{x})$. Общий вид функционала для наилучшей оценки $\lambda(\mathbf{x})$ нам неизвестен. Указанные в [2, 4] частные решения для λ зависят от выбранных критериев оптимальности и имеют недостатки. Определим некоторые требования к свойствам наилучшей оценки $\lambda(\mathbf{x})$.

1. Коэффициент $\lambda(\mathbf{x})$ должен оцениваться в заданной точке $\mathbf{x}$ с качеством комбинированной оценки регрессии не хуже, чем с оценками (4).
2. Коэффициент $\lambda(\mathbf{x})$ должен оцениваться для конечного объема выборочных данных N и не терять качество оценки вплоть до нескольких десятков наблюдений.
3. В условиях конечного объема выборки N выражение оценки $\lambda(\mathbf{x})$ должно быть *применимо* для *любой* выбранной точки $\mathbf{x}$. То есть свободно от ситуаций неопределенности (или сингулярности).
4. При решении задачи оценки регрессии частой ситуацией является отсутствие какой либо дополнительной информации, кроме выборочных данных (конечного объема N). Поэтому в качестве критериев оптимизации для адаптивной оценки $\lambda(\mathbf{x})$ должны быть взяты выборочные критерии.
5. Комбинированная оценка регрессии с наилучшей оценкой $\lambda(\mathbf{x})$ должна по возможности минимизировать СКО оцениваемой величины $J(\mathbf{x})$, а не выборочные регрессионные остатки.

Одним из подходов к решению задачи может быть новое определение вида функционала для представления оценки $\lambda(\mathbf{x})$, отличающееся от вида (4). Необходимо также определить состав аргументов для этого функционала, критерии и параметры его оптимизации. Представим коэффициент $\lambda(\mathbf{x})$ в виде

$$\lambda(\mathbf{x}) = \varphi(\mathbf{a}(\mathbf{x})) + \xi, \tag{5}$$

где $\varphi()$ – некоторый функционал, представляющий выражение коэффициента $\lambda(\mathbf{x})$; $\mathbf{a}$ $(\mathbf{x})$ – вектор аргументов, которые необходимо оценить в точке $\mathbf{x}$; ξ – величина, учитывающая ошибки представления и оптимизации коэффициента $\lambda(\mathbf{x})$.

Рассмотрим еще один способ оценки коэффициента λ, удовлетворяющий указанным требованиям и выражению (5).

Статистическая оценка коэффициента λ, зависящая от выбора точки x

Представим аппроксимацию неизвестного выражения для коэффициента $\lambda(\mathbf{x})$ с помощью функционалов, подобных ядерным оценкам. Обозначим их $k(x)$. Выберем функцию ядра следующего вида, рекомендованную в [6]

$$k(x) = \frac{1}{1+|x|^p},$$

где x – аргумент функции; p – величина, принимающая значения в виде целых положительных чисел.

Выбранная функция дает удовлетворительное приближение оцениваемой величины при удалении от области, содержащей наблюдения, и удовлетворяет требованию 3.

Для вектора $\mathbf{a}(\mathbf{x})$ можно предложить несколько вариантов получения различных аргументов в точке $\mathbf{x}$. Упростим задачу выбора и ограничимся дополнительной информацией о некоторых функционалах плотности в точке $\mathbf{x}$, полученных на основе выборочных данных и предложенных в [5] для оценки коэффициента $\lambda(\mathbf{x})$ из другого критерия. Выберем для формирования аргументов следующие величины [2, 5]:

$$C(\mathbf{x}) = g_2(\mathbf{x}) - J_N(\mathbf{x})g_1(\mathbf{x}),$$
$$V(\mathbf{x}) = C(\mathbf{x}) + (J_N(\mathbf{x}))^2 - J_N(\mathbf{x})g_1(\mathbf{x}),$$

$$\Delta_0(\mathbf{x}) = g_1(\mathbf{x}) - J(\mathbf{x};\boldsymbol{\theta})g_0(\mathbf{x}),$$

$$\Delta_1(\mathbf{x}) = \Delta_0(\mathbf{x})\left(1 - \frac{1}{1+(NH)^{\frac{3}{4}}(\Delta_0(\mathbf{x}))^2}\right), \quad H = \prod_{j=1}^{m} h_N^{(j)},$$

$g_0(\mathbf{x})$, $g_1(\mathbf{x})$, $g_2(\mathbf{x})$ имеют вид:

$$g_0(\mathbf{x}) = \frac{1}{NH}\sum_{i=1}^{N}\prod_{j=1}^{m} K[(x^{(j)} - X_i^{(j)})/h_N^{(j)}],$$

$$g_1(\mathbf{x}) = \frac{1}{NH}\sum_{i=1}^{N} Y_i \prod_{j=1}^{m} K[(x^{(j)} - X_i^{(j)})/h_N^{(j)}],$$

$$g_2(\mathbf{x}) = \frac{1}{NH}\sum_{i=1}^{N} Y_i^2 \prod_{j=1}^{m} K[(x^{(j)} - X_i^{(j)})/h_N^{(j)}].$$

Здесь $K(u)$ – заданное ядро (некоторая функция плотности вероятности); $h_N^{(j)}$ – параметры масштаба, удовлетворяющие следующим требованиям: при $N\to\infty$, $h_N^{(j)}\to 0$, $NH\to\infty$.

Обозначим еще одно выражение оценки функционала плотности в точке **x**

$$a_0(\mathbf{x}) = (J_N(\mathbf{x}))^2 - J_N(\mathbf{x})g_1(\mathbf{x}).$$

Тогда $V(\mathbf{x})$ можно представить как:

$$V(\mathbf{x}) = C(\mathbf{x}) + a_0(\mathbf{x}).$$

Выбранные аргументы являются ненормированными величинами, что может вызвать проблемы их применения в выражении (5). Проведем нормировку выбранных аргументов, взяв их отношения, и определим вектор аргументов в следующем виде

$$a_1(\mathbf{x}) = \frac{a_0(\mathbf{x})}{C(\mathbf{x})}, \quad a_2(\mathbf{x}) = \frac{V(\mathbf{x})}{NH(\Delta_1(\mathbf{x}))^2}, \qquad \mathbf{a}(\mathbf{x}) = \{a_1(\mathbf{x}), a_2(\mathbf{x})\}. \tag{6}$$

Аргументы в виде (6) не имеют размерности и соответствуют требованию 3.

Для аргументов (6) необходимо подобрать параметры масштаба $h_N^{(j)}$ некоторым наилучшим способом на основе выбранного критерия оптимальности. Если выбор критерия затруднен, сделаем предположение, что параметры масштаба, выбранные для оценок непараметрической модели регрессии $J_N(\mathbf{x})$, дадут удовлетворяющие результаты и при использовании их в аргументах выражения (6).

И последнее замечание. Для сохранения вида зависимости (прямая или обратная) коэффициента λ по отношению к оценкам регрессии из выражения (2), в оценке $\lambda(\mathbf{x})$ будем использовать функцию $(1-k(x))$. Для компенсации потерь оптимальности оценки $\lambda(\mathbf{a}(\mathbf{x}))$, в том числе от «чужих» параметров масштаба, применим в ней коэффициенты сглаживания. С учетом выше изложенного представим адаптивную оценку $\lambda(\mathbf{x})$ в виде

$$\hat{\lambda}(\mathbf{a}(\mathbf{x}),\mathbf{b}) = b_3\hat{\lambda}(\mathbf{a}(\mathbf{x}), b_1, b_2) = = b_3\prod_{j=1}^{2}(1 - k(a_j(\mathbf{x})b_j)), \tag{7}$$

где $\mathbf{b}=\{b_1,b_2,b_3\}$ – вектор коэффициентов сглаживания, значения которых подберем из следующего выборочного критерия оптимальности

$$Q_2(\mathbf{b}) = \sum_{i=1}^{N}[Y_i - J_N(\mathbf{X}_i;\hat{\lambda}(\mathbf{a}(\mathbf{x}),\mathbf{b}))]^2 \to \min_{\mathbf{b}}. \tag{8}$$

Сократим количество подбираемых элементов вектора **b**, применив способ, предложенный в [4]. Зафиксируем значения коэффициентов b_1, b_2, а b_3 оценим по выборке. Представляя выражение для b_3, введем обозначения

$$\hat{\lambda}(\mathbf{a}(\mathbf{X}_i), b_1, b_2) = \hat{\lambda}(\mathbf{a}(\mathbf{X}_i); b_1,\dots,b_3 \mid b_3 = 1),$$
$$d_1(\mathbf{X}_i) = Y_i - J_N(\mathbf{X}_i),$$
$$d_2(\mathbf{X}_i) = J_N(\mathbf{X}_i) - J_N(\mathbf{X}_i;\hat{\boldsymbol{\theta}}). \tag{9}$$

Используя выражения (2), (7), (8), (9), получим выборочную оценку для b_3

$$b_3 = \frac{\sum_{i=1}^{N} d_1(\mathbf{X}_i)d_2(\mathbf{X}_i)\hat{\lambda}(\mathbf{a}(\mathbf{X}_i), b_1, b_2)}{\sum_{i=1}^{N}(d_2(\mathbf{X}_i)\hat{\lambda}(\mathbf{a}(\mathbf{X}_i), b_1, b_2))^2}.$$

Оценка (7) удовлетворяет большинству требований постановки задачи.

Модифицированная оценка коэффициента λ, зависящая от выбора точки x

Можно ли улучшить оценку вида (7), используя еще какую либо дополнительную информацию на основе выборочных данных? Предположим, что одна из построенных моделей регрессии (параметрическая или непараметрическая) получает прогнозные оценки, близкие к истинным значениям. Тогда сумма выборочных регрессионных остатков от ее оценок может быть меньше, чем у другой модели. Используем это предположение. Обозначим суммы квадратов выборочных регрессионных остатков построенных параметрической и непараметрической моделей регрессии:

$$Q_3 = \sum_{i=1}^{N}[Y_i - J(\mathbf{X}_i;\hat{\boldsymbol{\theta}})]^2, \quad Q_4 = \sum_{i=1}^{N}[Y_i - J_N(\mathbf{X}_i)]^2.$$

Взяв их отношение, получим безразмерную нормированную величину $s=Q_4/Q_3$.

Величину s можно использовать разными способами. Предложим один из них. Сформируем дополнительный аргумент в виде некоторой функции, учитывающей предпочтение оценок одной из построенных моделей (величина s), сближение оценок построенных моделей при увеличении объема выборки (N), а также, вида зависимости коэффициента λ: $a_3=f(s,N)$. Используем его в выражении (6) как величину, увеличивающую предпочтение оценок одной из построенных моделей. Тогда вектор аргументов должен включать a_3:

$$\mathbf{a}_1(\mathbf{x}) = \{a_1(\mathbf{x}), a_2(\mathbf{x}), a_3\}. \tag{10}$$

Проверим свойства комбинированных оценок регрессии вида (2) с оценками коэффициентов λ вида (4), (7) и с вектором (10) в имитационном эксперименте.

Выбор вида оценок моделей регрессии для эксперимента

Для оценок параметрической модели регрессии выберем оценки линейной параметрической модели $J(\mathbf{x};\boldsymbol{\theta})=\theta_0+\theta_1x^{(1)}+...+\theta_mx^{(m)}$ с оценкой вектора $\boldsymbol{\theta}=(\theta_0,..., \theta_m)$ по методу наименьших квадратов. Для оценок непараметрической модели регрессии выберем оценки вида Надарая–Ватсона [7]

$$J_N(x^{(1)},...,x^{(m)}) = \frac{\sum_{i=1}^{N} Y_i \prod_{j=1}^{m} K[(x^{(j)} - X_i^{(j)}) / h_N^{(j)}]}{\sum_{i=1}^{N} \prod_{j=1}^{m} K[(x^{(j)} - X_i^{(j)}) / h_N^{(j)}]}, \quad (11)$$

где $K(u)$ – заданное ядро (некоторая функция плотности вероятности); $h_N^{(j)}=c_j(m/N)^{(1/(4+m))}$ – параметры масштаба; c_j – выбираются из условия минимума суммы квадратов регрессионных остатков

$$Q(c_1,...,c_m) = \sum_{i=1}^{N} [Y_i - J_N(X_i^{(1)},..., X_i^{(m)})]^2 \to \min_{c_1,...,c_m}.$$

Условия имитационного эксперимента

Сравнение качества оценок представленных моделей регрессии – непростая задача, так как оценки получены с использованием разных критериев оптимальности, разными методами, с несколькими допущениями. Поэтому имитационный эксперимент проведен с общим критерием качества оценок регрессии – оценки СКО.

Для простоты интерпретации результатов имитационный эксперимент проведен с одномерной величиной $x\in R^1$. Сравнение свойств оценок выполнено со случайным выбором точек X_i. Для имитации случайных ошибок измерений величины X и Y генерировались из двумерного нормального закона с заданным коэффициентом корреляции. Параметры генерации выборок обеспечивали в эксперименте формирование функции «истинной» регрессии – линейной $J(x)=0,0+1,0x$.

Для имитационного эксперимента выбраны следующие условия моделирования:

- сравнение качества оценок в условиях малых объемов выборки выполнено для объемов выборок в диапазоне N=5–125 с интервалом δ=5;
- случайные величины Y и X из двумерного нормального закона генерировались из условий нулевых математических ожиданий, со среднеквадратическим отклонением $\sigma(X)=\sigma(Y)=1,0$ и коэффициентом корреляции r=0,7;
- для каждого значения N числовые результаты эксперимента были получены по серии исходных выборок количеством K=1350000 (с одинаковым объемом наблюдений N в каждой выборке).

Оценки непараметрической модели регрессии для каждого наблюдения выборки получены по методу скользящего эксперимента (с исключением из выборки наблюдения i в точке X_i, в которой оценивается регрессия).

Сравнение качества оценок регрессии проводилось у пяти моделей ($J_1=J(X;\hat{\boldsymbol{\theta}})$, $J_2=J_N(X)$, $J_3=J_N(X;\hat{\lambda}_o)$, $J_4=J_N(X;\hat{\lambda}(\mathbf{a}(X)))$, $J_5=J_N(X;\hat{\lambda}(\mathbf{a}_1(X)))$). Для выявления способностей комбинированной оценки уменьшать СКО оценки истинной регрессии, использовались следующие критерии. Критерии, вычисленные для каждой исходной выборки.

1. Средняя квадратичная погрешность прогнозирования в произвольных точках

$$S_j(J) = \frac{1}{N}\sum_{i=1}^{N}[J(X_i) - J_j(X_i)]^2, \quad j = \overline{1,5}, \quad (12)$$

где $J(X_i)$ – истинное значение регрессии в произвольной точке X_i; $J_j(X_i)$ – оценка регрессии по модели j в произвольной точке X_i.

Критерии, вычисленные по серии K исходных выборок.

2. Среднее квадратичных погрешностей прогнозирования по серии K в произвольных точках

$$Q_j(J) = \frac{1}{K}\sum_{s=1}^{K} S_j(J), \quad j = \overline{1,5}. \quad (13)$$

При указанных условиях моделирования в имитационном эксперименте обеспечивается преимущество (в смысле качества) оценкам параметрической модели регрессии $J(X;\hat{\boldsymbol{\theta}})$.

Результаты эксперимента

Результаты имитационного эксперимента по критериям (12) и (13) представлены на рисунке. Номера кривых на рисунке совпадают с номерами моделей регрессии, указанных в условиях эксперимента. Кривые на графиках рисунка статистически значимо различаются на уровне значимости меньше 0,05, за исключением областей пересечения кривых.

На рисунке представлены результаты сравнения качества оценок регрессии по критерию $Q_j(J)$. По горизонтальной оси указан объем выборки N, по вертикальной оси – величина среднего квадратичных погрешностей прогнозирования по серии K в произвольных точках.

По результатам сравнения при указанных условиях имитационного эксперимента можно сделать следующие выводы.

1. Оценки модели регрессии:
   - параметрической J_1 дают третий по качеству результат среди представленных моделей в случае истинной линейной регрессии, проигрывая оценкам комбинированных моделей J_3 и J_5 (рисунок, кривая 1);
   - непараметрической J_2 дают невысокий результат среди представленных моделей (рисунок, кривая 2). Но с увеличением объема выборки N качество оценок улучшается;

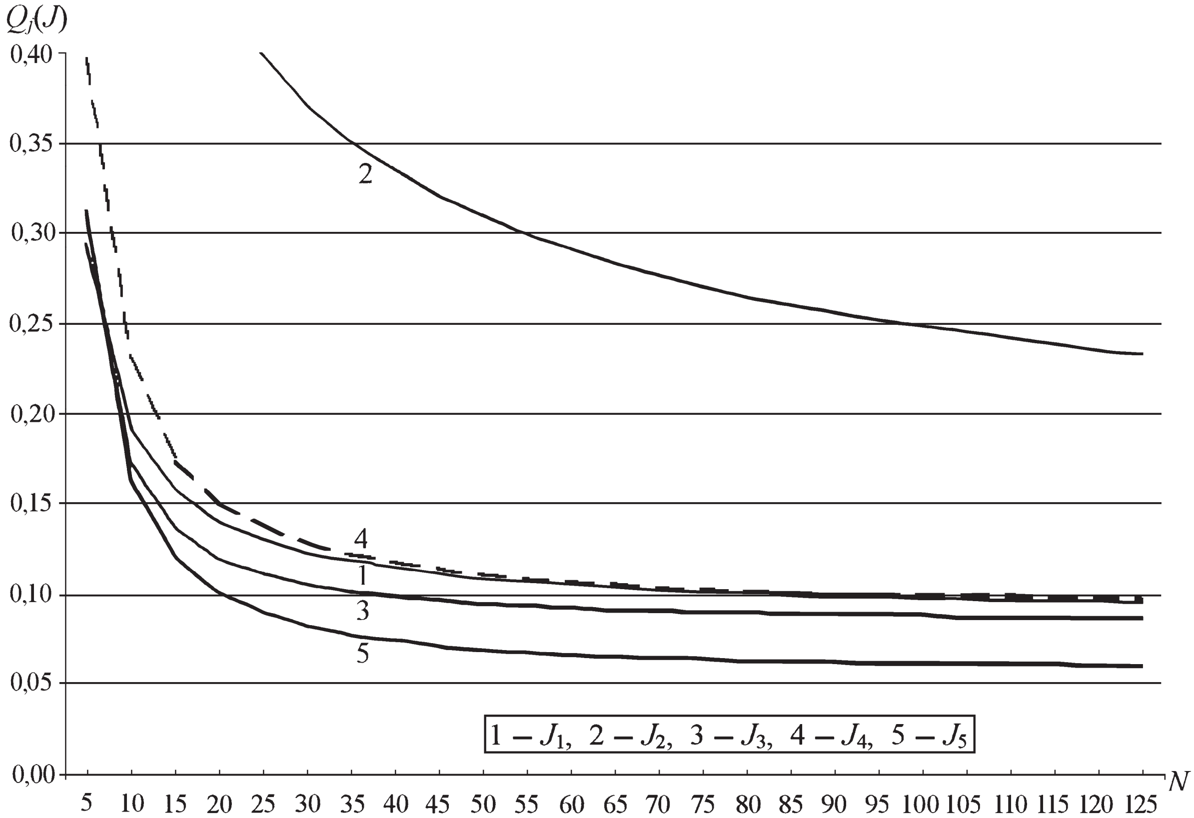

Рисунок. *Среднее квадратичных погрешностей прогнозирования по серии K в произвольных точках $Q_j(J)$ при коэффициенте корреляции r=0,7*

- комбинированной J_3 дают второй по качеству результат (рисунок, кривая 3), проигрывая лишь оценкам модели J_5. Это подтверждает заявленные свойства оценок комбинированной модели регрессии (2);
- комбинированной J_4 дают результат лучше, чем у непараметрической модели J_2 (рисунок, пунктирная кривая 4), а при объемах выборки N>35 практически не отличаются от оценок параметрической J_1 (и на уровне значимости меньше 0,05);
- комбинированной J_5 дают лучший результат среди всех моделей (рисунок, кривая 5). С увеличением объема выборки N выигрыш по сравнению с оценками других моделей сохраняется. Это также подтверждает заявленные свойства оценок комбинированной модели регрессии (2).

2. Оценки весового коэффициента $\lambda(X)$ с вектором вида (10) в комбинированной оценке регрессии J_5 показали свои преимущества по сравнению с оценками вида (7) и (4). Они удовлетворяют всем требованиям постановки задачи. Сложность вычисления оценок с вектором (10) почти не отличается от вида (7).
3. В эксперименте подтверждены предположения, что дополнительная информация вида (10) является достаточной для оценки коэффициента λ в точке (**x**) и используется эффективнее, чем в оценках вида (7) и (4), несмотря на неоптимальную оценку параметров масштаба $h_N^{(j)}$ для оценок (6).
4. Результаты эксперимента позволяют утверждать, что оценки $\lambda(X)$ с вектором вида (10) могут стать основой для дальнейшего улучшения комбинированных оценок регрессии в точке **x** и достижения наилучшего качества оценок.

Заключение

В работе представлена статистическая оценка весового коэффициента λ в точке **x** вида (7) и с вектором вида (10) для комбинированной модели регрессии вида (2) и продемонстрирована ее работоспособность при конечных объемах выборок N.

В имитационном эксперименте подтверждены свойства комбинированной модели регрессии вида (2) получать оценки лучше, чем у каждой из построенных моделей (параметрической или непараметрической). Результаты имитационного эксперимента демонстрируют преимущества ее оценок уже при N>10.

Лучшие оценки комбинированной модели J_5 и хорошие оценки модели J_3 в имитационном эксперименте подтверждают близость оценок коэффициентов λ к наилучшим. Это подтверждает правильность выбора критериев оптимизации и величин дополнительной информации (6) и (10), а также функционалов вида (5) для оценки λ.

Представленные варианты комбинированных оценок регрессии и результаты имитационного эксперимента позволяют утверждать, что у комбинированных оценок имеются широкие возможности совершенствования.

Результаты эксперимента показывают, что представленная комбинированная оценка регрессии со статистической оценкой весового коэффициента λ предпочтительнее, чем каждая из оценок параметрической и непараметрической моделей. Их преимущества проявляются как при малых объемах выборок, так и при их интерполяции за пределы эксперимента (N>125).

Результаты моделирования, представленные на рисунке, получены с помощью кластера Межрегионального вычислительного центра ТГУ СКИФ Cyberia (skif.tsu.ru). Автор выражает благодарность сотрудникам Центра за оказанную помощь.

СПИСОК ЛИТЕРАТУРЫ

1. Дмитриев Ю.Г. Непараметрическое условное оценивание функционалов плотности распределения // Математическое моделирование и теория вероятностей / под ред. И.А. Александрова, А.М. Бубенчикова, В.Н. Берцуна, Ю.К. Устинова. – Томск: Изд-во «Пеленг», 1998. – С. 169–177.
2. Скрипин С.В. Свойства комбинированной оценки регрессии при конечных объемах выборок // Известия Томского политехнического университета. – 2008. – Т. 313. – № 5. – С. 10–14.
3. Скрипин С.В. Свойства комбинированной оценки в задаче классификации наблюдений // Известия Томского политехнического университета. – 2009. – Т. 314. – № 5. – С. 21–26.
4. Скрипин С.В. О весовом коэффициенте в комбинированной оценке регрессии // Новые информационные технологии в исследовании сложных структур: Тезисы докл. VIII Росс. конф. с международным участием. – Томск, 5–8 октября 2010. – Томск: Изд-во НТЛ, 2010. – С. 98–99.
5. Скрипин С.В. Комбинированная непараметрическая оценка регрессии // Вестник Томского государственного университета. – 2005. – Прил. № 14. – С. 311–313.
6. Дукарский О.М., Левит Б.Я. Некоторые применения непараметрических оценок регрессии // Многомерный статистический анализ в социально-экономических исследованиях / Академия наук СССР, Центральный экономико-математический институт. Серия «Ученые записки по статистике». – М.: Наука, 1974. – Т. XXVI. – С. 31–37.
7. Надарая Э.А. Об оценке регрессии // Теория вероятностей и ее применения. – 1964. – Т. 9. – Вып. 1. – С. 147–149.

Поступила 12.04.2011 г.

УДК 65.012.122

ОПТИМИЗАЦИЯ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С ГИСТЕРЕЗИСНОЙ СТРАТЕГИЕЙ УПРАВЛЕНИЯ ОДНОТИПНЫМ РЕЗЕРВНЫМ ПРИБОРОМ

Л.И. Самочернова, Е.С. Петров

Томский политехнический университет
E-mail: evst@tpu.ru

Исследована система массового обслуживания с гистерезисной стратегией управления однотипным резервным прибором, управляемым по текущему времени ожидания заявки, находящейся первой в очереди. Проведена оптимизация системы при учете потерь на ожидание и амортизацию.

Ключевые слова:
Система, обслуживание, время ожидания, амортизация, оптимальный момент.
Key words:
System, service, queuing time, depreciation, optimal moment.

Введение

В последние годы сфера приложений теории массового обслуживания расширилась благодаря бурному развитию вычислительных, телекоммуникационных и производственных систем. Управляемые системы массового обслуживания (УСМО) являются адекватными моделями многих реальных технических систем [1–9]. Значительное число работ посвящено изучению систем массового обслуживания (СМО), в которых моменты включения и выключения резервных приборов, интенсивность обслуживания зависят от числа заявок в системе или от длины очереди [1–4]. Однако существуют лишь отдельные работы, например [5–9], в которых изучены СМО с управлением по времени ожидания. В данной статье рассматривается УСМО, которая аналогична системе, изученной в [8], но с гистерезисной стратегией управления однотипным резервным прибором.

1.Описание системы

Рассмотрим однолинейную систему массового обслуживания с простейшим входящим потоком интенсивности λ, к которой может подключаться ре-